

DOCUMENT RESUME

ED 200 648

TM 810 263

AUTHOR Edel, Robert L.
TITLE Some Advantages of Alternate-Choice Test Items.
PUB DATE Apr 81
NOTE 8p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Los Angeles, CA, April 11-17, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Comparative Testing; Difficulty Level; Guessing (Tests); *Multiple Choice Tests; *Objective Tests; *Test Format; *Test Items; *Test Reliability
IDENTIFIERS Item Discrimination (Tests); Kuder Richardson Reliability Formulas

ABSTRACT

An alternate-choice test item is a simple declarative sentence, one portion of which is given with two different wordings. For example, "Foundations like Ford and Carnegie tend to be (1) eager (2) hesitant to support innovative solutions to educational problems." The examinee's task is to choose the alternative that makes the sentence most nearly true. There is no element of verbal knowledge, no mental ability that cannot be tested through the use of alternate-choice items. These items are relatively simple to write and to use. They are not limited to testing recall of factual details. In a study comparing a test of alternate-choice items with an equivalent test of true-false items, 28 students enrolled in an educational measurement course were rewarded for taking both tests by using only the higher of the two test scores to determine their final grades. Statistical analysis of the results from these tests showed that the tests composed of alternate-choice items were somewhat less difficult and considerably more reliable than were the true-false tests. (Author/RL)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED200648

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY A.L. Ebel

TM 8/0 263

Some Advantages of Alternate-Choice Test Items

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

1. The nature of alternate-choice items

An alternate choice test item is a simple declarative sentence, one portion of which offers two different versions of an idea. The examinee's task is to choose the version that makes the sentence most nearly true. Here are some examples.

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC),

1. The resistance by teachers to the use of teaching machines is motivated mainly by fear of losing 1)a job 2)a sense of self-worth.
2. Published comments about education today are mainly 1)complimentary 2)critical.
3. If a child is achieving poorly in school, the parents are most likely to blame 1)the child's inherited intelligence 2)the school's instructional program.

In many situations there are good reasons for favoring the use of such alternate-choice items. Like all choice-type items, they tend to be more objective, efficient and reliable than supply-type items. They are as efficient as true-false items and, as shown in this study, they tend to yield more reliable scores.

2. Alternate-choice vs. multiple-choice items

Alternate-choice items have some important advantages over the more familiar multiple-choice item form that offers four answer choices. They are more efficient in that they yield more scorable responses per unit of testing time. They are easier to write, because they require only two alternative answers.

Often the important questions an item writer would like to ask have only two plausible alternative answers. A problem is major or minor, simple or complex. Action in response to it is reasonable or unreasonable. The President either supports or opposes a restriction on the import of foreign automobiles. The birth rate in Russia is higher or lower than that in the United States. Increasing the homogeneity of the items in a test either increases or decreases the reliability of the test scores. Polar alternatives like these are very common in real life.

Test constructors who are required to write four alternative items have developed a number of ways for combining alternate-choice ideas into four-choice questions. They may ask, in fact or in effect, which one of the following statements is true or untrue. They may combine two questions, asking both for a judgment and a justification. They may offer two statements and ask the examinee whether A only, B only, both, or neither is true. Each of these preserves the four-choice form at the cost of the more detailed information on the examinee's ability, and the more reliable measurements of that ability, that alternate-choice items would provide.

3. The problem of guessing

The general preference for four-alternative multiple-choice items over two-alternative items probably reflects concern over the possibility of correct response from blind guessing. But fears of the harmful effects of guessing tend to be greatly exaggerated. Although much blind guessing is possible, little of it is done actually because it promises much less reward than thoughtful response. The certain effect of blind guessing is to lower the reliability of the test scores. Since tests composed of alternate-choice items often yield scores of very acceptable reliability, as they did in this study, it is clear that guessing need not limit the usefulness of such items.

Some interesting data on this point were provided by a study of multiple-choice vocabulary test items with different numbers of answer choices.¹ In tests composed of equal numbers of items, those with four choices, as expected, produced more reliable scores than those offering three alternatives. Those with three yielded more reliable scores than those with two. But two choice items can be

¹Robert L. Ebel and Bob J. Williams. "The effect of varying the number of alternatives per item on multiple-choice vocabulary test items". 14th Yearbook, National Council on Measurements Used in Education, pp. 63-65, 1957.

answered more quickly than three, and three more quickly than four. When tests of equal duration are compared the order of the different forms in reliability was reversed. Two-choice items produced the most reliable scores, and four-choice items the least.

4. Writing true-false items

The findings of this study supported the use of true-false (i.e. two-choice) items,² and led to efforts to improve their effectiveness.^{3,4} One of the suggestions developed in this effort was that true-false items should always be conceived in pairs. For every true statement there should be a parallel false statement, and vice versa. For example:

1. Temperatures change less from season to season on the equator than in the arctic. (T)

Temperatures change more from season to season on the equator than in the arctic. (F)

These item pairs suggested another possibility. Suppose the two were combined and presented as one alternate-choice item like this:

1. Temperatures change 1)less 2)more from season to season
on the equator than in the arctic.

How would a test composed of such items compare with an equivalent test in true-false form?

5. A comparison of true-false and alternate-choice items

An answer to this question was sought from a class of 28 students enrolled in a course in educational measurement. The course content was divided into eight instructional units. A 25-item test was given at the conclusion of each unit.

²Robert L. Ebel. "The Case for True-False Test Items". School Review. 78:373-389, May 1970.

³Robert L. Ebel. "How to Write True-False Test Items". Educational and Psychological Measurement. 31:417-26, Summer 1971.

⁴Robert L. Ebel. "Can Teachers Write Good True-False Test Items?". Journal of Educational Measurement. 12:31-35, Spring 1975.

In previous terms these tests had been composed entirely of true-false items. For purposes of the present study the class agreed to take two 25-item tests on each unit. One of these was the usual true-false test. The other was an alternate-choice test. The two tests were composed of independent samples of items. They did not ask the same questions. The reward for taking two tests was that only the higher of the two test scores would be used to determine the student's final grade.

A comparison of the results from these tests is presented in Table I.

TABLE I

Statistics for True-False and Alternate-Choice Unit Tests

Unit	Mean		Stand.Dev.		KR20		Difficulty		Discrimination	
	TF	AC	TF	AC	TF	AC	TF	AC	TF	AC
1	18.28	20.52	3.59	3.05	.71	.58	27%	18%	.36	.29
2	19.12	21.08	3.05	2.81	.63	.64	24%	16%	.26	.27
3	18.61	22.30	2.13	2.58	.13	.69	26%	11%	.21	.22
4	19.90	18.90	3.77	3.36	.78	.68	20%	24%	.36	.34
5	20.96	21.09	1.84	2.37	.33	.64	16%	16%	.18	.22
6	16.46	18.58	2.40	3.83	.22	.76	34%	26%	.24	.36
7	19.13	20.70	2.24	2.60	.28	.56	23%	17%	.22	.25
8	17.93	16.74	3.49	4.02	.68	.74	28%	33%	.33	.40
Mean	18.80	19.99	2.81	3.08	.47	.67	25%	20%	.28	.30

Note that the tests composed of alternate-choice items were somewhat less difficult and considerably more reliable on the whole than were the true-false tests. A reasonable hypothesis to explain these results is that the alternate-choice form reduces some of the ambiguity inherent in true-false items.

6. The "limitation" of testing for knowledge

What of the limitations of alternate-choice items? While it is true that they are limited to testing for knowledge, this is not a serious limitation. Knowledge is power. While other things beside knowledge, things such as effort, personality and luck contribute to success in most human enterprises, the contribution of knowledge is paramount. To help pupils develop solid and extensive structures of knowledge, along with decent characters, is about all that any teacher or any school can do to help them live successful, happy lives.

The knowledge we are talking about is not information stored in memory. It is information that has been digested and assimilated as part of a person's structure of beliefs. Knowledge is produced out of information by thinking. If the information is not understood, it does not become knowledge. To understand an item of information, one must see its relation to other elements of a structure of knowledge. Like knowing, understanding is a matter of degree. It is possible to know something without fully understanding it. But in that case, one does not know it fully either. As the terms are used here "knowing" and "understanding" are practically synonymous.

Any mental ability consists entirely of knowledge. To do some mental task a person needs only to know completely how to do it. To say that the person may lack the ability to do what he knows perfectly well how to do is unreasonable. And if a person is unable to use the knowledge possessed it must simply be for lack of sufficiently well-developed knowledge of the right kind. It cannot be because knowledge itself is inadequate. Hence, the limitation of two-alternative items to testing knowledge is in reality no limitation at all.

7. Are alternate-choice items too simple?

The apparent simplicity of alternate-choice items has led some to suppose that they are suitable only for testing recall of trivial factual details. Not so. Ideas of great scope and importance can be expressed in simple declarative sentences.

1. The apparent motion of the sun is due to rotation of the earth.
2. The properties of the chemical elements are periodic functions of their atomic numbers.
3. Complex organisms evolved from simpler organisms.
4. All men are mortal.

If an alternate-choice item requires only recall of a trivial factual detail, the fault lies not in the form but in the formulator.

Surely there are complex problems and there are simple problems. But complex problems consist entirely of a sequence of simpler problems. This is true of a game of chess or of a heart transplant. If each simple problem is solved correctly, the game is won and the operation succeeds. The whole is precisely equal to the sum of its parts. The ability to do the whole job is no higher than that required to do any of the parts.

Some test constructors seem to be inordinately fond of complex test items, perhaps under the impression that complex items provide better indications of an examinee's "real" ability, that they require application of "higher mental processes". There is superficial plausibility to such beliefs, but no good evidence to support them. We know too little about mental processes at any level to justify assertions about "real" abilities or "higher" mental processes.

8. The limitations of complex test items

What we do know for sure, and have good evidence for believing, is that complex test items tend to be inefficient, and to yield unreliable scores. Complex items include many words, and thus require much time for reading, reflection and response.

This means that fewer items can be recorded on a test page, or included in a test of specified time limits. That means, in turn, that the scores are likely to be much less reliable. An important basic principle of test construction is that the larger the number of independent scorable responses per unit of time, the higher the reliability of the scores is likely to be.

When an examinee considers a complex test item many thoughts are likely to go through his or her head; true ideas, false ideas, relevant ideas, irrelevant ideas. After all that thinking only one response is recorded to be scored correct or incorrect. If the contributory thoughts were examined and scored independently would not one obtain a much more extensive and detailed assessment of the examinee's knowledge? Alternate-choice items enable an examiner to do just that.

Another reason why complex test items tend to yield unreliable scores inheres in their very complexity. There are so many more chances for the examinee to go astray, misperceiving this or overlooking that. These chance elements may have little to do with the examinee's overall ability. They make the test items more difficult without making them better measures of achievement in learning.

9. The case for alternate-choice items

There is no element of verbal knowledge, no mental ability, no achievement in cognitive learning that cannot be tested by using alternate-choice items. They are not limited to testing for recall of factual details. They yield scores of very respectable reliability. They are relatively simple to write and to use. Whatever can be tested using more complex item forms can be tested more efficiently and more accurately using alternate-choice items. They deserve wide use.